Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-04-05 14:04:15
REFERENCE NO: 246

Author Names & Affiliations

- Timothy Lance - NYSERNet

Contact Email Address (for NSF use only)

(Hidden)

Research Domain, discipline, and sub-discipline

Mathematics (geometric topology, complex analysis), joint work in biology, physics, atmospheric sciences, and economics. As NYSERNet president for 18 years and now as CRO, working with advanced networking and high performance computing.

Title of Submission

Reaching Across Boundaries to Shape Future Cyberinfrastructure

Abstract (maximum ~200 words).

NYSERNet provides the research network and other network related services in New York, but even more we are a community of trust that has convened researchers, computing and networking experts, academic administration, industry, and the government to discuss how to work together on problems like energy, climate, health care, and economics that cross disciplinary, institutional, and sector boundaries. These ongoing discussions seem to remarkably parallel the questions in NSF 17-031, and the response below comes out of our efforts over many years toward the goals enunciated in this NSF solicitation. The responses are at a fairly high level, not focusing on high energy physics, or genomics, or health care. But our efforts over the years have revealed remarkable congruence of needs even if they differ in ancillary details. Given the high level nature of this response, we welcome any request to drill down in any particular area.

Thank you.

Question 1 Research Challenge(s) (maximum ~1200 words): Describe current or emerging science or engineering research challenge(s), providing context in terms of recent research activities and standing questions in the field.

New York has an extraordinarily rich and diverse research community involving major research universities, corporations (research headquarters for IBM, GE, Corning, and many others), government (State labs, Brookhaven National Labs) and private research institutions (Cold Spring Harbor, Trudeau and others). Areas of research encompass an incredibly broad spread: genomics, fundamental research on disease process, high energy physics, LIGO, computational chemistry, research across a spectrum of biological processes from transport across a cell membrane to brain architecture and function, and efforts to mimic the brain in CMOS. Many fundamental discoveries will only come by reaching across disciplinary, institutional, and sector boundaries. Though the wet lab components of these efforts remain, the data generated is growing rapidly, and attention to transport, storage, and computational of that data steadily grows in importance.

To frame the research challenges, we first look at several collaborative efforts and the ways the resulting achievements elucidate both the opportunities and challenges for the work enabled by these collaborations. NYSERNet, one of the most enduring partnerships in the Internet, is one such collaboration. Created by the research and education community, for more than three decades NYSERNet has deployed ever more advanced purpose built research networks for its community. In the perilous years following the attack on 9/11, NYSERNet deployed fiber in New York City, created what has become one of the most important R&E exchange points in 32 Avenue of the Americas in lower Manhattan, set up a statewide optical network, and built a 9/11 inspired business continuity center in Syracuse that proved its mettle during hurricane Sandy.

All of this required and deepened the trust within the community. With network resources in place and computers rapidly becoming more powerful, we could imagine taking on harder problems. In January, 2008 a group of researchers, HPC experts and representative from research administration from academe, industry and government gathered at a NYSERNet hosted conference the New York Academy of Sciences to discuss how to work together on problems like health care, energy, and climate that cross disciplinary, institutional, and sector boundaries. The most critical need identified was for “bridging people” with enough disciplinary background and technical expertise, whether it be in computing, instrumentation, software, or other areas, that they could enable research groups to make productive use of new technologies. One immediate consequence was the creation of the high performance computing partnership HPC-NY (https://hpc-ny.org/) between academic institutions and industry. NSF spoke to us at length about this long before initiatives such as ACI-REF.

Just the process of trying to do this, even if we failed, expanded the fabric of trust, and this was the case in high performance computing, genomics, weather and climate change, health care and a host of other areas. The efforts to grow and deepen the partnerships that already exist and expand into new areas are ongoing. Much of this, driven by research of academic institutions and industry, happens without government support. These are careful gambles into speculative areas – e.g. neuromorphic computing – which, as they mature, we hope grant programs will follow.

We identify a number of research challenges which, while not discipline specific, in our experience seem to seem to be universal across disciplinary boundaries.

Sharing Data and Expertise: This has many components, including identifying relevant data and researchers where this is most appropriate, building a level of trust where researchers are willing to consider it, and making the case for the benefit of sharing. Extreme events sometime enable this. By the grace of God NYSERNet’s PoP at 60 Hudson Street in NYC kept power, our SONET ring in the City stayed up as well as the statewide backbone, and part of the rapid recover of Internet in NYC, Westchester, and Long Island stemmed from use of our facilities. Partnerships with the commercial sector formed then remain strong today. Following Sandy, researchers uptown whose labs were spared Sandy’s wrath opened their doors to research competitors downtown, and many acknowledge insights that depended critically on sharing. HPC-NY is an example of a collaboration not formed under duress that endures. All three components need to be done together, all depend on trust.

Intellectual Property: In New York this has been an ongoing effort between academic institutions and industry, an effort of critical importance since the research benefits from the partnerships are so great. We submit that there is no easy answer but will need constant attention. A joint venture may begin where the initial motivation is the potential for discovery, but then part of the outcome becomes valuable. Occasionally clearly identified rules apply that cleanly decide who accrues benefit. As often as not the development happens in an unanticipated direction and without trust and good will a fair resolution is much more difficult. The current unfinished debate over CRISPR demonstrates how hard this can be.

Data and Idea Replication: This is a very important subject today and, unfortunately, the conclusion that seems to be drawn about an experiment that cannot be replicated, or a concept that cannon be proved, is that lack of integrity or carefulness by the originator. That is rarely the case. This is discussed at length in question 3.

Exploring New Areas of Research: Deciding what to pursue and how to fund it is one of the toughest choices. For individual researchers it becomes a choice by them and their collaborators of which problems seem the most central, or perhaps for which do they have the best set of conceptual and technical tools to begin the attack. At a more senior level, research administrators sometimes have to decide among equally worthy areas requiring more funds than are available. And occasionally the health of an institution depends on strategic institution-wide decisions (like pursuing genomics, or brain inspired computing).

Our view is that, particularly in new areas, access to publically available funding should be should come later. That does not mean that such funding is not critical. The LIGO discovery would never have happened without the National Science Foundation. But the ideas came first. Half a century passed after Einstein’s original geometric formulation of gravity before the research community convinced itself that gravitational waves were real, not an artifact of the formulation.

Investing in the New Talent: Universities, industry, and the National Science Foundation (as well as NIH and others) all have a critical role here, as do their support organizations like NYSERNet, which invested in technology to support a set of talented LIGO researchers. The resulting dynamic wave network could be used by all, but was inspired by the critical work of talented LIGO researchers. We are concerned at the difficulty younger researchers face in getting their first grant support. As discretionary money dwindles as a share of the federal budget this problem will only get worse.

Question 2 Cyberinfrastructure Needed to Address the Research Challenge(s) (maximum ~1200 words): Describe any limitations or absence of existing cyberinfrastructure, and/or specific technical advancements in cyberinfrastructure (e.g. advanced computing, data infrastructure, software infrastructure, applications, networking, cybersecurity), that must be addressed to accomplish the identified research challenge(s).

Much of the collaborative effort in New York over the last decade and a half grew out of the discussions hosted by the NYSERNet community at the New York Academy of Sciences, a roughly concurrent discussion on “Data Driven Research” by the late Dr. Robert Richardson, Nobel Laureate, of Cornell University, and growing realization and appreciation of the mutual dependence of academe and industry on each other. But if, as we believe, our research is increasingly data driven, then a robust research network is an absolute necessity. We feel fortunate to have a large enough research community in the state to be able to afford a dedicated, purpose built research infrastructure.

We have also been fortunate to have had resources to act quickly when opportunities arose and grant resources were not available. Support for the deployment of fiber in New York City and creation of what became the global exchange point came entirely from internal funds. No federal or state money was used. The same is largely true of our subsequent work largely because we had the resources, the NYSERNet Board identified them as critical needs for them, and authorized the expenditure. The same practice continues today in, for example, our replacement of optics, leveraging the remarkable recent changes.

But nationally the research enterprise depends on the health of every regional network. Many have missions that have been shaped by a variety of external influences. They may be an agency of the state, or have a purpose set by a board of regents, or have pursued a path mostly focused on supporting K-12, and a host of other reasons. Some face real challenges. We refer the reader to the excellent Quilt response to this RFI for a national overview of the regional networks. Their health is critical to the sustainability of the national R&E transport infrastructure.

Continuing to pursue the thread of data driven research we must address not just the transport of data but its manipulation and storage. As the data generated continues to grow very rapidly, computing and storage must be just as nimble and agile as the networks to be equal to the demands of the research. The collaboration HPC-NY very consciously includes industry, and we have all benefitted. The tapestry is more intricate, however. Major computing resources are housed at individual institutions, sometimes affordable by agreements with companies like IBM and with costs partially defrayed by grants. For many institutions this is a delicate balance between concurrent demands of those contributing, whether by grants or contributions, and sorting out intellectual property issues between contributors and institutions. This must be weighed against a judgement of whether far more can be achieved by the partnership than could be done alone.

Our general experience is that partnerships between universities and between universities and industry always yield more, and as a community we willingly take on the burden of sorting through intellectual property and other issues. Typically we have found that the rules related to such partnerships are far more flexible and less prescriptive than those that typically come with grants, often mandates handed to agencies like NSF or NIH. But to the extent possible, we urge that regulations coming from government funding allow as much flexibility as possible so that wonderful outcomes of pursuing “what if … “ aren’t precluded.

We continue to sort through the right balance of cloud services versus in house resources for computing and storage. Some very tightly couple computations are still best achieved with complete control of the computing environment. However, the cloud as a component of what we use gives far greater flexibility as we sort through fundamental questions of where data should be stored and processed or even, given enough control, of basic computational architecture to best support the research.

In parallel with all of this, many university researchers work, with university permission, with companies like IBM, GE, Google, Corning, and many others, under NDA, with intellectual property issues a negotiation between the institutions and companies. Faculty also work with some perhaps unexpected companies like Proctor and Gamble and Kellogg with hard CFD problems that have impacted ordinary day to day life. Tools honed for studying turbulence and clotting in blood flow or storm development and intensification carry over.

Long a concern for both academic and corporate members of the NYSERNet community, security has gained even greater prominence of late for a host of reasons. The cost of producing some research data has been significant, with LHC and LIGO as prime but certainly not the only examples. Ransomware, an unfortunate new term in our lexicon, has largely been solved by more frequent copying and distribution. Some data has become extremely valuable. Senseless though it may be, our work can be interrupted by entities who do not necessarily want to steal, just disrupt. Virtually every institution in the NYSERNet community has a chief security officer, typically reporting to the president and independent of IT. Security teams at multiple institutions share, regularly communicate with technical staff at NYSERNet, and participate in REN-ISAC to be better able to anticipate and respond to attacks and breaches. Institutions are working with corporate partners, typically under NDA, as new tools are developed.

Security is an area where catalytic support from NSF could provide great benefit. The amount of data continues to grow rapidly, and we are yet to discover the security issues related to the shipping, manipulating, and storing of data sets orders of magnitude greater than anything we have seen before.

The integration and reliability of cyberinfrastructure tools, data replication, and dealing with the rapid development of these tools, closely interwoven and the most important challenge before us, are dealt with separately in question 3.

Question 3 Other considerations (maximum ~1200 words, optional): Any other relevant aspects, such as organization, process, learning and workforce development, access, and sustainability, that need to be addressed; or any other issues that NSF should consider.

We list here several issues and concerns drawn from the first two questions. Some of these are areas where the issue lies principally with NSF or where NSF can take the lead, and others where this can be a joint effort with the community it serves, and some which fall to each of us not in our organizational roles but as citizens.

Workforce development: Workforce development is a spectrum of issues depending on the age and circumstance of the person who would benefit. At the most advanced level we are concerned especially about the difficulty young researchers face in securing support for their research, especially the first grant. In Question 1 we discussed the critical need for “bridging people”, a conclusion reached in a joint meeting of academe, industry, and government, a conclusion reaffirmed several times since and, we now know, by others as well. Unfortunately, such people do not have an academic home so the positions are not sustainable as yet. NSF grants could include a small amount of support for such positions, and a request that the institution sustain them in some way – university money, or perhaps state workforce development support.

At the youngest end of the age spectrum, we are losing children not just from STEM, but from education generally, with heavy losses in middle school, with minorities hardest hit. Many city school districts report minority graduation rates below 10%. The responsibility falls on all of us as citizens. The author of this document has several decades experience working in area schools, sometime with wonderful results, but successful programs were difficult to scale, labor intensive, and need to deal with ever greater poverty and inequity.

Sustainability: In Question 1 we emphasize the need for healthy regional networks and refer to the Quilt’s submission to this RFI. NSF probably does not have a role here. In networking and computing grants over the past several decades, and in particular with the various CC* solicitations, NSF has provided critical seed money to enable networking or computing expansions/enhancements. NSF appropriately asserted that sustaining what it enabled was the responsibility of the grantee institution. But sustainability of existing cyberinfrastructure remains a concern, especially as budget tightens which the National Science Foundation should continue to monitor.

Replicability, and Integration and Reliability of the Cyberinfrastructure Stack: We believe these seemingly independent issues are in fact tightly coupled. Replicability of data and or experiments is a hot button issue these days. The long delay between the first LIGO chirp on September 14, 2015 and the announcement the next year was to give the entire global team time to check and recheck the instruments, data, programming, and every other technical aspect to be sure this momentous discovery was no fluke. Subsequent detections, and the extraordinary strength of the signal above background noise, also helped.

But often when research results cannot be replicated the impression left by those reporting it was tampering with data. Unfortunate incidents like East Anglia, as well as lack of public awareness of the turmoil necessary before new ideas are firmly understood, feed into this. We believe that many factors are vastly more likely behind non-replicable data or experiments. These include:
1. Changes in computing. Earlier work may have been done on machines where greater precision was unaffordable, or where memory limitations forced the time step in, say, atmospheric modeling, to be greater.
2. Changes in instrumentation. Older data may rely on less sensitive instrumentation. In genomics, availability of many high throughput sequences allow multiple runs and greater convergence that was possible even a few years ago.
3. Software errors. Any programs of even moderate complexity has some unintended behaviors. We have evidence of this even in well-known open source programs.
4. Miscommunication between machines and software, or between software and software, or between researchers and the technology tools. Again we have example of this, in some cases where the error was discovered indirectly and then very subtle and hard to find directly.
5. Errors from scaling or parallelization. As we submit larger jobs to bigger machines the code often does not parallelize as well as we expect. One of the errors in number 5 came from this.
6. Hidden biases. Parts of the cyberinfrastructure stack seem so automated (like screening out noise from possibly interesting signal at the level of a trigger farm) that we may not look for an unintentional hidden bias that will skew results.
7. Human error. No matter how careful we are, it happens. A fairly famous recent example came in the announcement of cosmic inflation from the South Pole data.

The various cyberinfrastructure tools on which we rely in general run independently of each other. Some attempt at integration is needed, and NSF could be the key catalyst, although we urge that any such attempts heavily involve industry. There are many parts to our cyberinfrastructure arsenal, all constantly being redeveloped and ideally improved, and industry is the driver for much of this. Thus all of the issues above will remain with us. Integration of all of this may not be possible, but the effort itself is worthwhile. Even if the components cannot be integrated, the people from disparate areas who will need to work together in the effort will themselves be better integrated, and the benefits from that will be long lasting.